DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

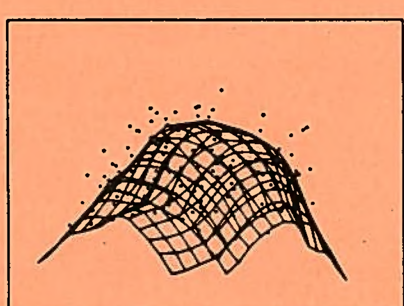
NONPARAMETRIC LIKELIHOOD RATIO INTERVALS

Art Owen

Technical Report No. 6

November 1985

Laboratory for Computational Statistics



Department of Statistics
Stanford University

This document and the material and data contained therein, was developed under sponsorship of the United States Government. Neither the United States nor the Department of Energy, nor the Office of Naval Research, nor the U.S. Army Research Office, nor the Leland Stanford Junior University, nor their employees, nor their respective contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any liability or responsibility for accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use will not infringe privately-owned rights. Mention of any product, its manufacturer, or suppliers shall not, nor is it intended to, imply approval, disapproval, or fitness for any particular use. A royalty-free, nonexclusive right to use and disseminate same for any purpose whatsoever, is expressly reserved to the United States and the University.

Nonparametric Likelihood Ratio Intervals*

by

Art Owen

Department of Statistics

Stanford University

L.C.S. Technical Report No. 6

December 1985

Abstract

The well-known nonparametric maximum likelihood estimate of a distribution function F_0 is the empirical c.d.f. $\hat{F}$. Point estimates for many statistical functionals $T(F)$ are taken to be $T(\hat{F})$. This article explores interval estimation of $T(F_0)$ by $(\inf_{R(F)\geq c} T(F), \sup_{R(F)\geq c} T(F))$, where R is a nonparametric likelihood ratio function for distributions. For the mean, under regularity conditions, it is shown that the interval has asymptotic coverage α where $-2\log c$ is $\chi^2_{(1,\alpha)}$, the α quantile of the chisquare distribution on 1 degree of freedom. Thus, the theorem of Wilks on the asymptotic distribution of the likelihood ratio has a nonparametric analog. From the mean the result extends to a family of M-estimators. The n.l.r. intervals are related to the bootstrap confidence intervals based on nonparametric tilting of Efron (1981, section 11).

* Work supported by an Office of Naval Research contract N00014-83-K-0472.

1. Introduction

Suppose the random variables $X_1, \ldots, X_n$ are an i.i.d. sample from a distribution F_0. The well-known non-parametric maximum likelihood estimate of F_0 is $\hat{F} = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, where δ_{X_i} is the distribution function of a point-mass at X_i.

The likelihood function evaluated at a distribution F with mass $w_i \geq 0$ on X_i is

$$L(F) = \prod_{i=1}^n w_i$$

and this is easily seen to be maximized at $\hat{F}$. In this article the function L is restricted to distributions with support only on the observed sample. That is, it is assumed that $\sum w_i = 1$. The likelihood ratio function

$$R(F) = L(F)/L(\hat{F}) = \prod_{i=1}^n nw_i$$

is defined on the same simplex as L. Throughout this paper F will be a generic member of the simplex identified with the generic weights w_i, $i = 1, \ldots, n$.

Suppose we are interested in estimating the functional T, defined on the simplex and on F_0. The nonparametric m.l.e. of T is $T(\hat{F})$. If T is continuous on the simplex, as many statistical functionals of practical interest are, then the image of $T(\cdot)$ over the part of the simplex in which $R(F) \geq c$ is an interval by the compactness of that subset. The smaller c is, the larger the interval is, and the family of intervals is nested. The idea of this paper is to use the intervals so obtained as a family of confidence intervals for T. It should be mentioned that the region $R \geq c$ is convex, since this fact could make the numerical optimization of T easier.

For fairly general parametric situations, Wilks (1938) showed that $-2\log(R)$ has an asymptotic chisquare distribution, where R is the ratio of the likelihood evaluated at the true parameter to the maximum of the likelihood function. This fact can be used to form confidence intervals for the parameter that have asymptotically correct coverage. The errors are of order $1/\sqrt{n}$. In some situations small sample distributions of the likelihood ratio are available. Thus in the parametric likelihood ratio interval problem there is a reasonable way to chose c.

In the completely nonparametric situation we cannot possibly be so lucky. The distribu-

tion theory for a likelihood ratio when the number of parameters is tending to infinity at the same rate as the number of observations will be different. In particular, if the true distribution is continuous it will always have a 0 likelihood ratio. If the functional $T(F)$ is badly behaved, as is for example an indicator that is 1 iff F is continuous, or if T is the mean and F_0 is Cauchy, then this approach cannot possibly work. Fortunately, many statistical applications are much less perverse, and the nonparametric likelihood ratio method is not degenerate. It is usually the case that one can get T right without getting F_0 right.

This article focuses attention on the mean although extensions to certain M-estimates follow almost immediately.

In the next section, the nonparametric likelihood ratio interval estimates for the mean are derived. Section 3 computes the asymptotic probability of covering the true mean by such an interval. Theorem 1 in that section shows that the asymptotic coverage of the mean can be obtained from the $\chi^2_{(1)}$ distribution in the same way as in regular one-dimensional parametric families with parametric likelihoods. In particular it follows that asymptotic non-parametric likelihood ratio tests for the mean can be obtained. In section 4, the n.l.r. intervals are extended to a class of M-estimates. Section 5 reconsiders the restriction to distributions supported on the sample.

A similar approach was used by Thomas and Grunkemeier (1975) to get confidence intervals for the survival function in the presence of censoring. They provide a heuristic argument for the asymptotic chisquare distribution of their estimator. Their situation differs in that the extrema of R over all distributions are usually equivalent to the extrema over distributions with support on the observed (failure) times. The only exception is the somewhat degenerate case where one seeks an interval estimate of the probability of survival past a time b that does not lie between the smallest and largest observed failure time.

The motivation here was to find a way to compute nonparametric confidence intervals without having to know which bootstrap technique to choose. That problem can be acute when, for example, the bias-corrected percentile method and the percentile method choose intervals skewed in opposing directions. (Recent bootstrap confidence interval methods may overcome this difficulty. See Efron (1985).) In many parametric situations likelihood ratio intervals

automatically get the right skewness, and the thought was that nonparametric likelihood ratios might also.

Ironically the resulting technique is similar to another bootstrap confidence interval technique, namely the nonparametric tilting bootstrap confidence intervals of Efron (1981,section 11). There are two differences. One difference is that the family of intervals is based on weights (11.8) instead of (11.1) of Efron (1981). The other difference is that the chosen member of the family of intervals is taken from the asymptotic chisquare result of section 3 rather than from resampling from tilted distributions. For the mean this could be quite a large computational saving. For more complicated functionals, the numerical optimization could be more work than resampling.

2. N.l.r. interval estimates of the mean

The mean is written as a functional as

$$M(F) = \sum_{i=1}^{n} w_i X_i.$$

To obtain the n.l.r. interval corresponding to c one must maximize and minimize $M(F)$ subject to the constraints $\sum w_i = 1$, $\prod w_i \geq c$ and $w_i \geq 0$. In practice it suffices to use the simpler constraint $\prod w_i = c$. For most functionals of interest the product constraint is binding. Even if it is not, a typical application involves calculating the bounds for many c values and one could use the equality constrained extrema to compute the inequality constrained values.

Using Lagrange multipliers, let

$$G = \sum w_i X_i + \lambda_1 (1 - \sum w_i) + \lambda_2 (\log c - \sum \log(n w_i)).$$

Setting the partial derivatives of G with respect to w_i to zero yields

$$w_i = \frac{\lambda_2}{X_i - \lambda_1}.$$

λ_2 can be obtained as a normalizing constant. For a given set of weights the maximum of M is obtained when the largest weights are attached to the largest X_i, and the minimum obtains when the weights decrease with X_i. Therefore λ_1 cannot be in the interval $(X_{(1)}, X_{(n)})$, for then the w_i would not be monotone. This also guarantees that they must all be of the same

sign, that is, they must all be positive. For the minimum $M^{lo} = \sum w_i^{lo} X_i$ where

$$w_i^{lo} = \frac{(X_i - X_{(1)} - \eta^{lo})^{-1}}{\sum_j (X_j - X_{(1)} - \eta^{lo})^{-1}}$$

and for the maximum $M^{up} = \sum w_i^{up} X_i$ where

$$w_i^{up} = \frac{(X_{(n)} - X_i + \eta^{up})^{-1}}{\sum_j (X_{(n)} - X_j + \eta^{up})^{-1}}$$

for appropriate η^{lo} and η^{up}.

The η values can be found by a one dimensional zero-finding algorithm such as Newton's method. For c near 1, the η values are very large and positive and the weights are nearly equal. As c decreases the η values approach 0, and all the weight goes to the appropriate extreme observation.

The limiting values of the bounds as c approaches 0 are $X_{(1)}$ and $X_{(n)}$, the largest and smallest observed values. This means that no n.l.r. confidence interval for the mean will go outside the observed range of the data. This stands in contrast to parametric confidence intervals which may easily go outside the observed range. The very extreme confidence limits in the nonparametric case are obviously going to be suspect. It would be interesting to learn how close to the edge one can safely go. In the parametric case the extreme confidence limits will depend heavily on difficult to verify assumptions about the tails of the distributions in the model and may therefore be equally dubious.

Instead of calculating the η values for a prescribed c, it is more convenient to calculate c, M^{up} and M^{lo} over a grid of ηs. This allows the plotting of the log likelihood ratio as a function of x, and the limits for any desired c can be obtained via interpolation.

The following observations were generated from the standard normal distribution:

$$-2.2 \ -1.6 \ -1.2 \ -1.0 \ -0.2 \ -0.0 \ 0.2 \ 0.3 \ 1.0 \ 1.2,$$

the mean of the observations is -0.34 and the mean square is 1.24.

The nonparametric log likelihood ratio function for the mean based on these points is plotted in figure 1 as a solid line. The unconnected dots represent the true log likelihood ratio function under the assumption that the sample is from $N(\mu, 1)$.

Figure 1 Parametric and Nonparametric
Likelihood Ratio Functions for a Normal Mean

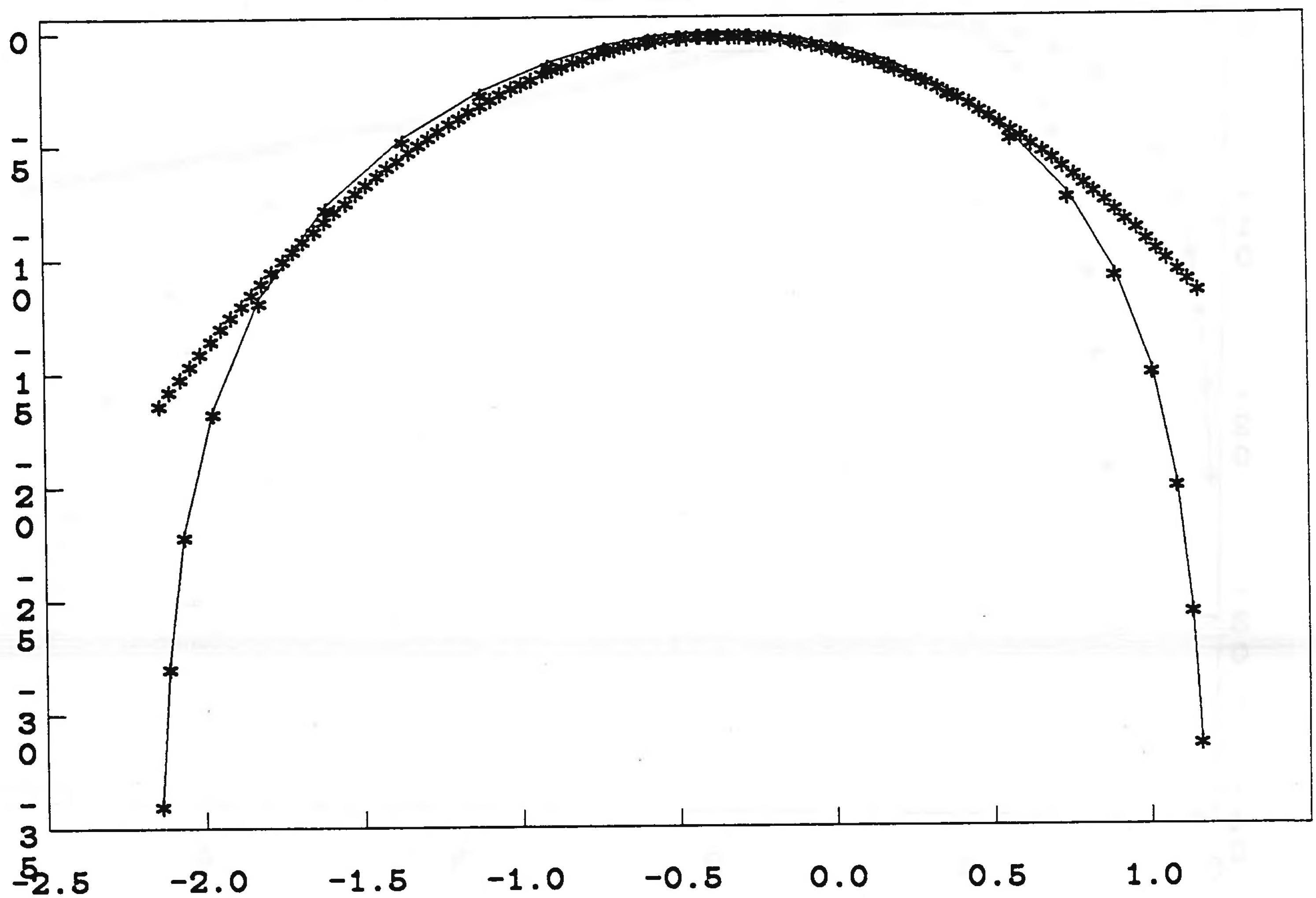
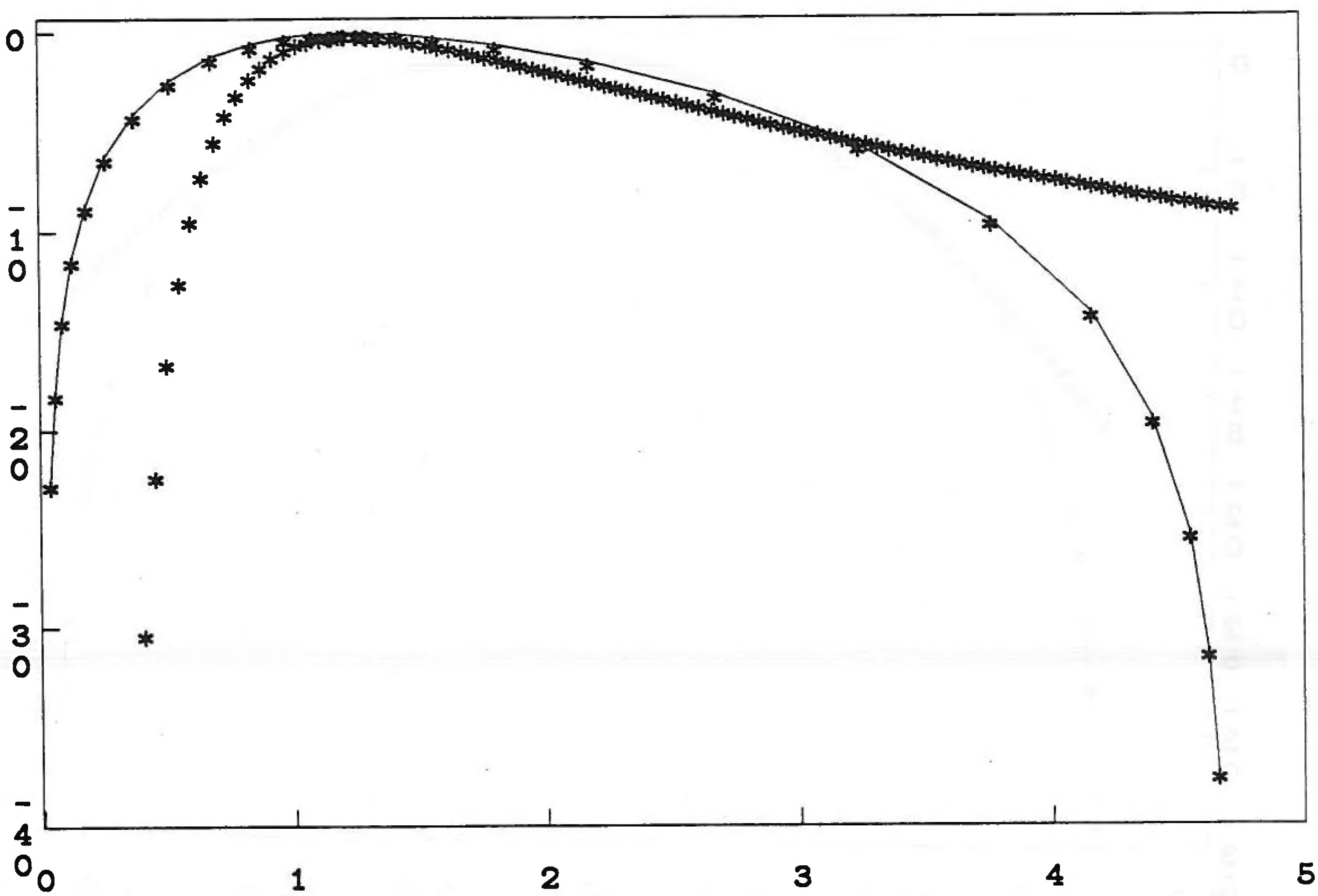


Figure 2 Parametric and Nonparametric

Likelihood Ratio Functions for a Normal Mean Square



Their squares are distributed as $\chi^2_{(1)}$, and the nonparametric log likelihood ratio function for the mean square is plotted in figure 2 as a solid curve. The unconnected dots in figure 2 represent the true log likelihood ratio function under the assumption that the sample is from $N(0, \sigma^2)$.

The nonparametric log likelihood ratio functions look surprisingly like the parametric ones, considering that only 10 points are available.

In figure 1 the n.l.r. interval endpoints were calculated for η values equal to $2^j(X_{(n)} - X_{(1)})$ for $j = -5, \ldots, 10$. In figure 2 the largest and smallest squared observations were used in the ηs.

The median also has a quite tractable algorithm for determining n.l.r. interval estimates, as do other quantiles but these reduce to certain quantities based on the binomial distribution.

3. Asymptotics for N.l.r. Estimation of the Mean

In this section it is proved that if F_0 has bounded support and is not degenerate that the n.l.r. interval for the mean of F_0 has asymptotic coverage α when $-2\log c = \chi^2_{(1,\alpha)}$, the α quantile of the chisquare distribution on 1 degree of freedom. The author conjectures that the bounded support condition can be replaced by existence of some moment higher than the first.

Theorem 1

Let $X_i \in [-M, M]$ be i.i.d. from a nondegenerate distribution F_0. If

$$X_U = \sup \textstyle\sum_{i=1}^n w_i X_i \quad \textit{and} \quad X_L = \inf \textstyle\sum_{i=1}^n w_i X_i$$

where both extrema are taken over $R(F) = \prod_{i=1}^n nw_i \geq c$, then

$$P(X_L \leq E(X_1) \leq X_U) \rightarrow \; P(\chi^2_{(1)} \leq -2\log c).$$

Proof: Without loss of generality $E(X_1) = 0$. $P(X_{(1)} < 0, X_{(n)} > 0) \rightarrow 1$, since F_0 is not degenerate, so we can assume that there is at least one observation of each sign. It follows that

$$R^\circ = \sup\{\textstyle\prod nw_i \mid \sum w_i = 1, \sum w_i X_i = 0\}$$

exists.

Now $X_L \le 0 \le X_U$ iff $c \le R^\circ$. We show that $-2\log R^\circ$ has an asymptotic chisquare distribution.

To find R°, let

$$G = \sum \log(nw_i) + \gamma(1 - \sum w_i) + n\lambda(0 - \sum w_i X_i).$$

Setting $\partial G/\partial w_i = 0$ one obtains

$$w_i = \frac{1}{\gamma + n\lambda X_i}$$

and summing $w_i \partial G/\partial w_i$ shows that $\gamma = n$. It follows that

$$\log(R^\circ) = \sum \log\left(\tfrac{n}{n+n\lambda^\circ X_i}\right) = -\sum \log(1 + \lambda^\circ X_i)$$

where λ° satisfies

$$0 = \frac{1}{n}\sum \tfrac{X_i}{1+\lambda^\circ X_i} \equiv g(\lambda^\circ). \tag{1}$$

The function $g(\lambda)$ defined by (1) is finite in an interval containing the origin. Since $g'(\lambda) = -\frac{1}{n}\sum_{i=1}^{n}(\frac{X_i}{1+\lambda X_i})^2 < 0$ (recall that not all $X_i = 0$), g is strictly decreasing in the interval and equation (1) is easily seen to have a unique solution in the interval $(-X_{(n)}^{-1}, -X_{(1)}^{-1})$. Note also that $g(0) = \overline{X}$, the sample mean.

λ° as defined by (1) is an M-estimate, although not one of location, with $\psi(X, \lambda) = X/(1 + X\lambda)$. The M-estimate applied to F_0 is 0 since F_0 has mean zero. Since $E(\psi(X, \lambda))$ is positive when $\lambda < 0$ and negative when $\lambda > 0$ it follows by proposition 3.2.1 of Huber (1981) that $\lambda^\circ \to 0$ strongly.

Expand $h = g^{-1}$ in a Taylor's series about $\overline{X}$, and evaluate it at 0 to get

$$h(0) = \lambda^\circ = h(\overline{X}) + (0 - \overline{X})h'(\overline{X}) + (0 - \overline{X})^2 h''(\xi)$$

where ξ is between 0 and $\overline{X}$. Now $h'(\overline{X}) = 1/g'(0)$ and $h''(\xi) = -g''(\eta)/g'(\eta)^3$ where $\eta = h(\xi)$ is between 0 and λ°. Therefore

$$\lambda^\circ = \frac{\overline{X}}{\overline{X^2}} + 2\overline{X}^2 \frac{\sum[\frac{X_i}{1+\eta X_i}]^3}{(\sum[\frac{X_i}{1+\eta X_i}]^2)^3}.$$

Let r° denote the second term in λ°. A simple bound is

$$|r^\circ| \le 2\overline{X}^2 n^{-2} \frac{1/n\sum|X_i|^3}{(1/n\sum X_i^2)^3} \frac{(1+|\eta|M)^6}{(1-|\eta|M)^3}.$$

Since $|\eta| < |\lambda^\circ| \to 0$ a.s. we can assume that $|\eta|$ is a small fraction of $1/M$. By the strong law of large numbers applied to both $|X_i|^3$ and X_i^2, the probability is arbitrarily close to 1 that $|r^\circ| \le 3\overline{X}^2 n^{-2} E(|X_1|^3)/E(X_1^2)^3 = O_p(n^{-3})$. By the central limit theorem $\sqrt{n}(\lambda^\circ - r^\circ)$ tends to the standard normal distribution and so $\lambda^\circ = O_p(n^{-1/2})$.

Now $P(\max|\lambda^\circ X_i| > .5) \to 0$ so we may use the same Taylor expansion of $\log(1+\lambda^\circ X_i)$ for all sufficiently large n:

$$\begin{aligned}
\log(R^\circ) &= -\sum \log(1+\lambda^\circ X_i) \\
&= -\sum(\lambda^\circ X_i - \tfrac{1}{2}(\lambda^\circ X_i)^2 + \eta_i) \quad \text{where } |\eta_i| < |\lambda^\circ X_i|^3/3 \\
&= -n\overline{X}\lambda^\circ + \frac{n}{2}\overline{X^2}\lambda^{\circ 2} - \sum \eta_i \\
&= -n\overline{X}\left(\frac{\overline{X}}{\overline{X^2}} + r^\circ\right) + \frac{n}{2}\overline{X^2}\left(\frac{\overline{X}}{\overline{X^2}} + r^\circ\right)^2 - \sum\eta_i \\
&= -n\frac{\overline{X}^2}{\overline{X^2}} - r^\circ n\overline{X} + \frac{n}{2}\frac{\overline{X}^2}{\overline{X^2}} + r^\circ n\overline{X} + \frac{n}{2}r^{\circ 2}\overline{X^2} - \sum\eta_i \\
&= -\frac{n}{2}\frac{\overline{X}^2}{\overline{X^2}} + \frac{n}{2}r^{\circ 2}\overline{X^2} + nO_p(n^{-3/2}) \\
&= -\frac{1}{2}(\chi^2_{(1)} + o_p(1)) + \frac{n}{2}O_p(n^{-6})O_p(1) + O_p(n^{-1/2}) \\
&= -\frac{1}{2}\chi^2_{(1)} + o_p(1)
\end{aligned}$$

so $-2\log R^\circ \to \chi^2_{(1)}$ in distribution as required.∎

The error in the approximation is of order $1/\sqrt{n}$ since the central limit theorem operates on the leading term at that rate and the other term is of that rate. This is the same rate obtained by Wilks (1938) for the parametric case.

4. N.l.r. intervals for M-estimates

The estimation of the n.l.r. confidence intervals for M-estimates is similar to that for the mean and the asymptotic chisquare distribution carries over too.

Let T be a statistical functional defined by

$$0 = \int \psi(X,T)dF(X)$$

where $\psi(X,\tau)$ is nondecreasing in X, continuous and nonincreasing in τ, and bounded. This covers most commonly used robust M-estimates of location; the main quibble from a practical point of view is that the concomitant estimation of the scale constant is being ignored.

Let $I_i(F) = dT(F)/dw_i$, the empirical influence of T at X_i. The upper bound of the n.l.r. interval for T for some c is

$$\sup_{R(F)\geq c} T(F).$$

The supremum is attained at weights satisfying

$$w_i = \frac{(I_i(F)-\lambda)^{-1}}{\sum(I_j(F)-\lambda)^{-1}}$$

by the same Langrangian multiplier method as was used in the case of the mean. The infimum is obtained similarly. As in the case of the mean the weights can be rewritten

$$w_i \propto (I_{(n)} - I_i(F) + \eta)^{-1}$$

for appropriate η where $I_{(n)}$ is the influence of $X_{(n)}$ and is also the largest of the influences.

The big difference here is that the $I_i(F)$ depend on the w_i through $F = \sum w_i \delta_{X_i}$, whereas for the mean $I_i(F) = X_i$ is fixed. Therefore even for a fixed η the corresponding point on the likelihood ratio curve will have to be estimated iteratively. Another approach is to obtain the likelihood ratio function for the M-estimate from that of the mean of $\psi(X,\tau)$ for a grid of values of τ. This idea underlies the proof of Theorem 2 below.

Most of the work of extending the asymptotic distribution from the mean to the M-estimates is done via the following lemma.

Lemma 1

Let $T(F)$ be the solution to $0 = \int \psi(X,T)dF(X)$ and let $\tau_0 = T(F_0)$. Assume that $\psi(X,\tau)$ is nondecreasing in X, and continuous and nonincreasing in τ, and that $T(F)$ exists for all F. Let $\tau^{up} = \sup_{R\geq c} T(F)$, let $M^{up} = \sup_{R\geq c}\sum w_i\psi(X_i,\tau_0)$ and define τ^{lo} and M^{lo} with infima replacing suprema. Then $\tau_0 \in (\tau^{lo},\tau^{up})$ implies $0\in[M^{lo},M^{up}]$ and $0\in(M^{lo},M^{up})$ implies $\tau_0\in(\tau^{lo},\tau^{up})$.

Proof: Suppose $\tau_0 < \tau^{up}$. Then each $\psi(X_i,\tau_0) \geq \psi(X_i,\tau^{up})$ and hence $\sup_{R\geq c}\sum w_i\psi(X_i,\tau_0) \geq \sup_{R\geq c}\sum w_i\psi(X_i,\tau^{up})$ i.e. $M^{up} \geq \sup_{R\geq c}\sum w_i\psi(X_i,\tau^{up})$. This supremum is nonnegative

since otherwise there would be weights w_i^* satisfying the likelihood ratio constraint for which $\sum w_i^* \psi(X_i, \tau^{up}) < 0$. But then $\tau' = T(\sum w_i^* \delta_{X_i})$ exists by the assumption on T and $\tau' > \tau^{up}$ by the continuity assumption, contradicting the definition of τ^{up}. Therefore $M^{up} \geq 0$. A similar argument shows that $\tau_0 > \tau^{lo} \Rightarrow M^{lo} \leq 0$, and so the first claim is proved.

Suppose $M^{up} > 0$. Then there are weights w_i^* satisfying the likelihood ratio constraint for which $\sum w_i^* \psi(X_i, \tau_0) > 0$. As above there exists a $\tau' > \tau_0$ for which $\sum w_i^* \psi(X_i, \tau') = 0$ and the definition of τ^{up} implies that $\tau^{up} \geq \tau' > \tau_0$. Similarly $M^{lo} < 0 \Rightarrow \tau^{lo} < \tau_0$. ∎

Now we can extend the asymptotic chisquare distribution to M-estimates via

Theorem 2

Let $T(F)$ be the solution of $\int \psi(X, T) dF(X) = 0$ be an M-estimate satisfying the conditions of Lemma 1. Suppose also that ψ is bounded. Let X_i be an i.i.d. sample of size n from F_0 with $\tau_0 = T(F_0)$. If $-2 \log c = \chi^2_{(1,\alpha)}$ then $P(\inf_{R \geq c} T(F) < \tau_0 < \sup_{R \geq c} T(F)) \to \alpha$.

Proof: By lemma 1

$$
\begin{aligned}
&P(\inf_{R\geq c} \sum w_i \psi(X_i, \tau_0) < 0 < \sup_{R \geq c} \sum w_i \psi(X_i, \tau_0)) \\
\leq &P(\inf_{R \geq c} T(F) < \tau_0 < \sup_{R \geq c} T(F)) \\
\leq &P(\inf_{R\geq c} \sum w_i \psi(X_i, \tau_0) < 0 < \sup_{R \geq c} \sum w_i \psi(X_i, \tau_0))
\end{aligned}
$$

and by Theorem 1 both of the bounding probabilities tend to α. ∎

5. Final remarks

The restriction to distribution functions with support on the observations is unnatural in a likelihood setting. We certainly do not believe that the true distribution is in the simplex determined by the sample. That n.l.r. intervals can work is just a sampling property of the random simplices we observe.

For the mean, the extrema that define the n.l.r. interval equal the extrema over all distributions with support on the closed interval $[X_{(1)}, X_{(n)}]$. This is easy to see, since if a distribution puts any probability on a set in the interval that contains no observations that probability can be "swept" to one or the other end producing a more extreme mean with no reduction in the likelihood ratio. Thus the extrema can be said to be taken over an uncountably infinite family of distributions. Without the restriction on the endpoints of the support,

weight ϵ can be put on a large unobserved X and give rise to an arbitrarily large mean with a likelihood ratio arbitrarily close to 1, and every n.l.r. interval becomes the whole real line.

The situation is different for M-estimators T with monotone ψ function and bounded influence (including the mean if we know a bound for X). Then it is easy to show that the constrained extrema of T over all distributions are not much different than those over the simplex and that the difference tends to zero as $n \to \infty$. A sketch of the proof is as follows. Consider the upper bound of the confidence interval. Any weight off the simplex might as well be put at ∞ (or at a known bound for X), so that the corresponding $\psi(\infty, \tau)$ equals the bound on ψ. The amount of weight w^* that can be put at ∞ while preserving $R \geq c > 0$ is bounded by $1 - c^{1/n} + c^{1/2}/n \to 0$. View this as contamination. The supremum of the contaminated M-estimate subject to the likelihood ratio condition is less than the uncontaminated supremum plus the greatest difference that the contamination could make at any underlying distribution. This latter quantity is $O(w^* B)$ where B is the bound on the influence function of T. It follows that n.l.r. intervals for T in which the limits are taken over all distributions on the line will have the same asymptotic coverage as those based on distributions supported on the sample.

Most bootstrap confidence interval techniques also work exclusively with distributions supported on the sample. An exception is Hjort (1985) who constructs a Bayesian bootstrap based on a Dirichlet prior.

References.

Efron, B. (1981). Nonparametric standard errors and confidence intervals. (with discussion) *Canadian Journal of Statistics*,**9** No. 2, pp. 139-172

Efron, B. (1985). Better bootstrap confidence intervals. L.C.S. Technical report 14, Department of Statistics, Stanford University, Stanford CA.

Hjort, N.L. (1985). Bayesian nonparametric bootstrap confidence intervals. L.C.S. Technical report 20, Department of Statistics, Stanford University, Stanford CA.

Huber, P.J. (1981). Robust Statistics John Wiley and Sons Inc., New York.

Thomas, D.R. and Grunkemeier, G.L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, **70**, pp. 865-871.

Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, **9**, pp. 60-62.